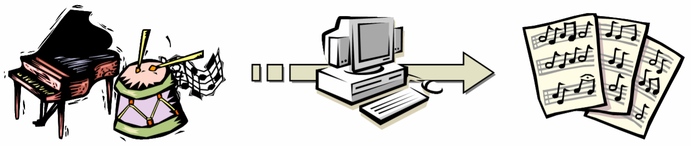
Automatic Transcription of Turkish Music



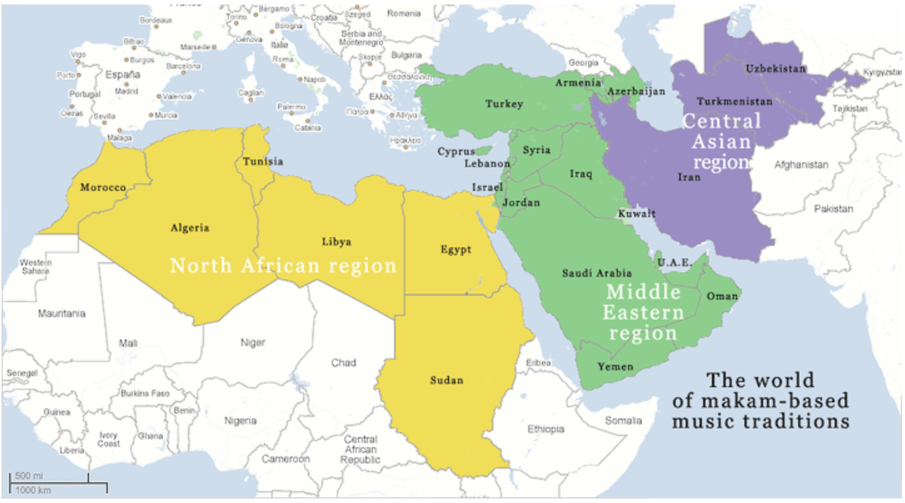
Music instrument sound　　Computer　　Score

By Barış Bozkurt,
Asst. Prof., Izmir Institute of Technology, Izmir/TURKEY

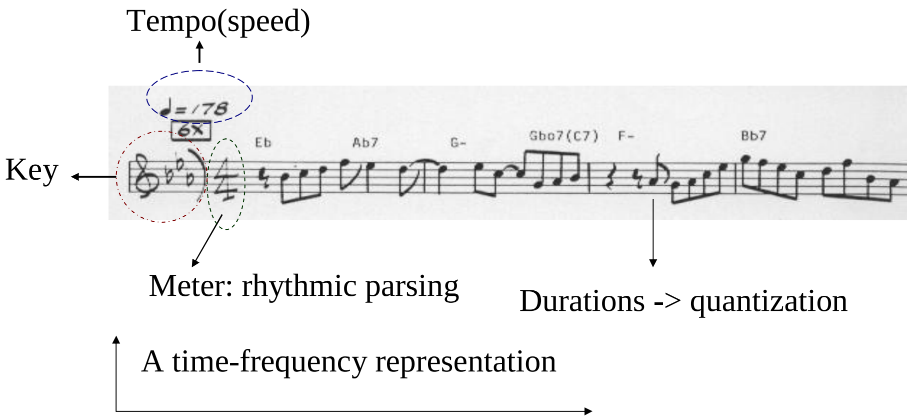
Ex-Researcher/PhD student, TCTS/FPMS,Mons/Belgium

Computational ethnomusicology, ethnic music processing:

an unbiased(non-12Tet) audio signal processing application domain



Automatic Music Transcription



Tempo(speed)

Key

Meter: rhythmic parsing

Durations -> quantization

A time-frequency representation

Basic problems:

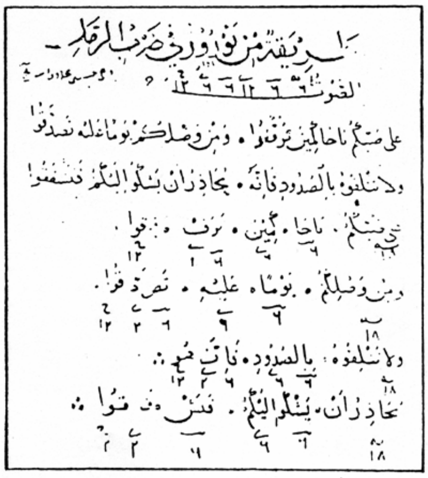
Pitch estimation and quantization

Duration estimation and quantization

Turkish Music Notation

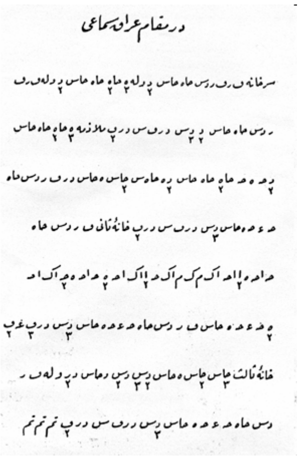
Ebced notation: 13th century

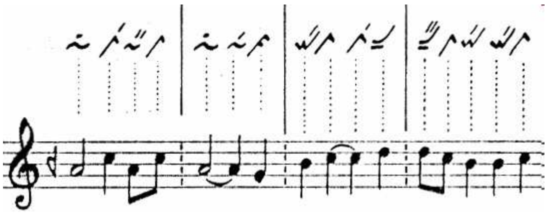
Re	Mi	Fa	Fa#	Sol
ا	د	ه	و	ح



Turkish Music Notation

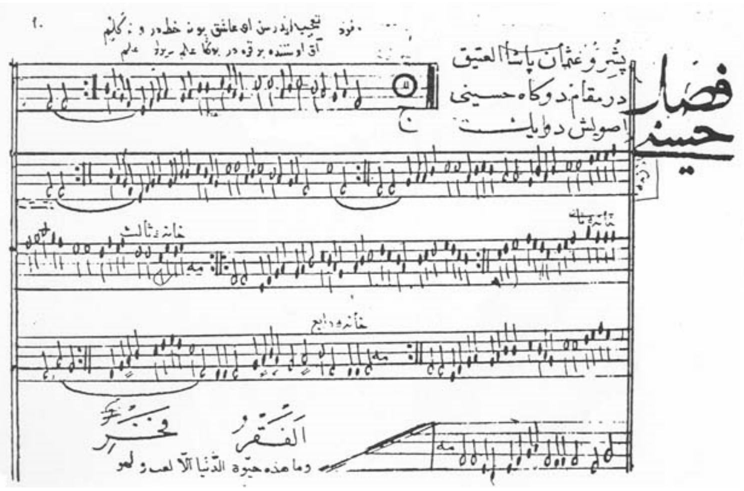
Hamparsum Limonciyan (1768-1839): Hamparsum notation





Turkish Music Notation

Albert Bobowski: a foreigner stayed in Topkapi Palace, uses Western notation for the first time(1650) to write Turkish music.

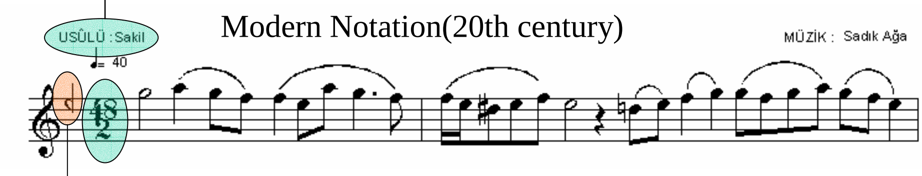


Automatic Transcription of Turkish Music

Extremely complex rhythmic structures: 9/8, 7/8, 13/8,...176/4

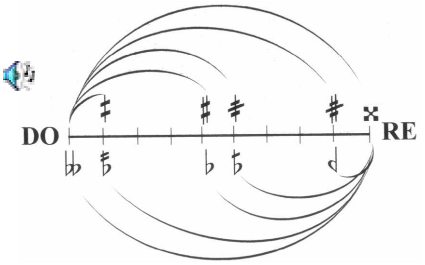
Ex: 3/2/2/2= 9/8

Modern Notation(20th century)



New accidentals since a minimum of 17 notes (debates continue) in an octave is needed.

Makam specifies the accidentals as the Key does in western music.

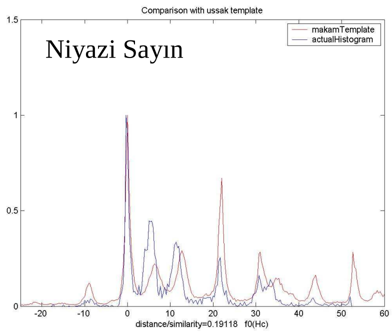


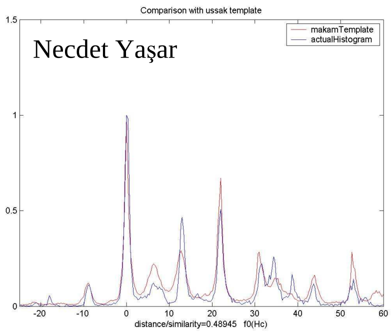
Differences from Western music that cause the need for modified/new algorithms

- Ahenk system: non-standard diapozon

- Freedom of the player, use of non-standard instruments

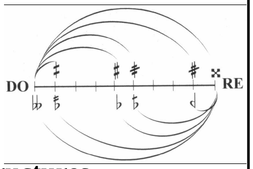
NEY TYPE / AHENK	*LENGTH (mm)*	*D^2 (mm)*	*A_4=440 cps*
Bolâhenk [i]	1014-1040	39-40	*neva*
Davud	925-936	35.5-36	*çargah*
Şah-Dik [ii] *Şah*	858-884	33-34	*segah – buselik*
Mansur	780-806	30-31	*dügah*
Kız	702-715	27-27.5	*rast*
Yıldız [iii]	650-663	25-25.5	*arak*
Müstahsen	598-611	23-23.5	*acem aşiran*
Süpürde	572-585	22-22.5	*aşiran*
Bolâhenk Nısfiye [iv]	520-533	20-20.5	*yegah*

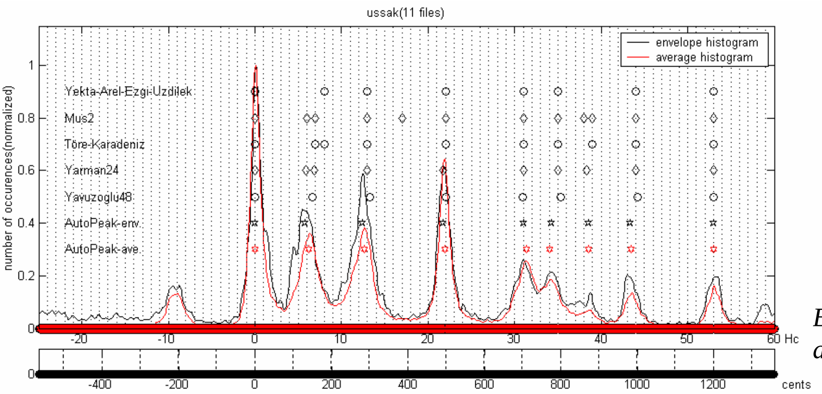




Differences from Western music that cause the need for modified/new algorithms

- More notes ... It is more close to micro-tonal music
- Theory is mostly descriptive, it is an oral tradition
  - **Mismatch of theory and practice**
  - **Highly complex makam/mode-key and usul/rhythm structures**

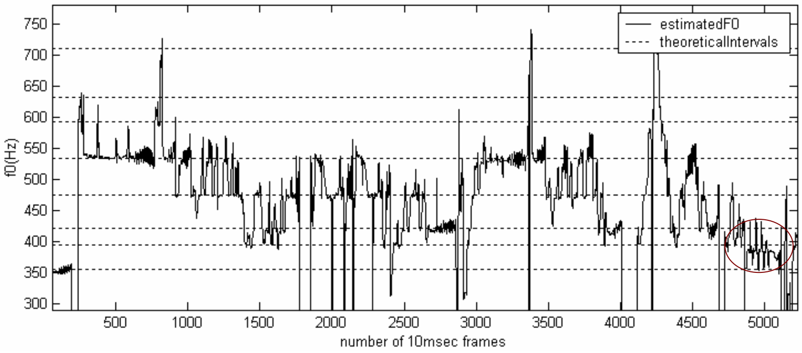




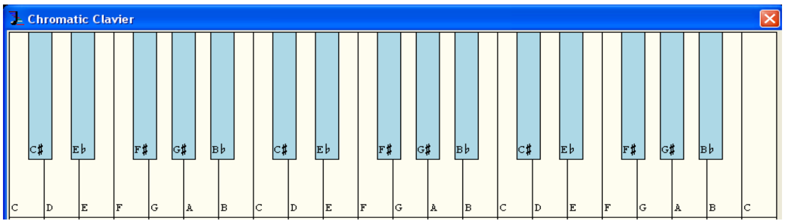
Bozkurt, et al, 2009

Differences from Western music that cause the need for modified/new algorithms

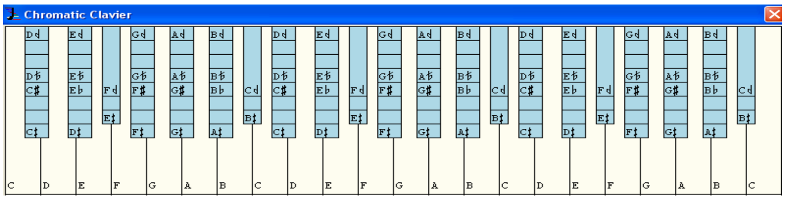
- Very frequent use of ornamentations



Comparing the two Tuning Systems



Western



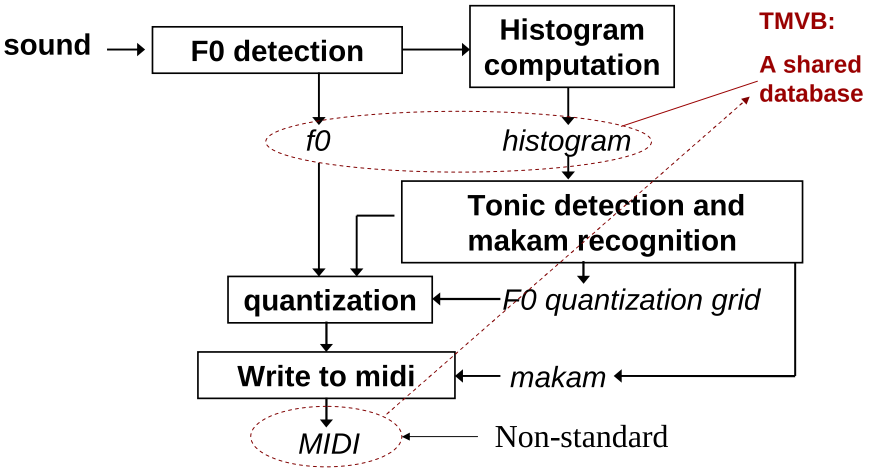
Turkish

OK ...there is also an important simplicity:

Monophonic/heterophonic -> half of automatic transcription literature is on multiple f0 estimation of polyphonic audio

---

Automatic Transcription of Turkish Music



sound

F0 detection

Histogram computation

TMVB:

A shared database

f0

histogram

Tonic detection and makam recognition

quantization

F0 quantization grid

Write to midi

makam

MIDI

Non-standard

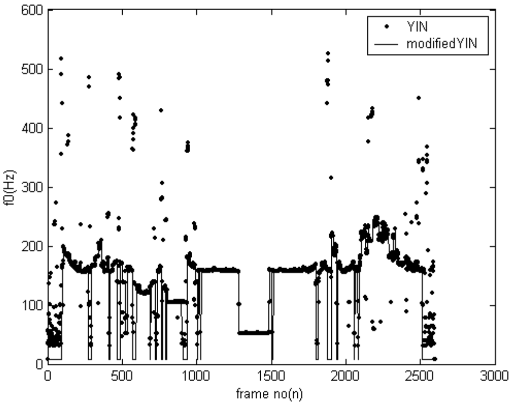
Rhythm analysis is excluded for now: our database consists of free rhythm solos: taksims.

F0 detection:

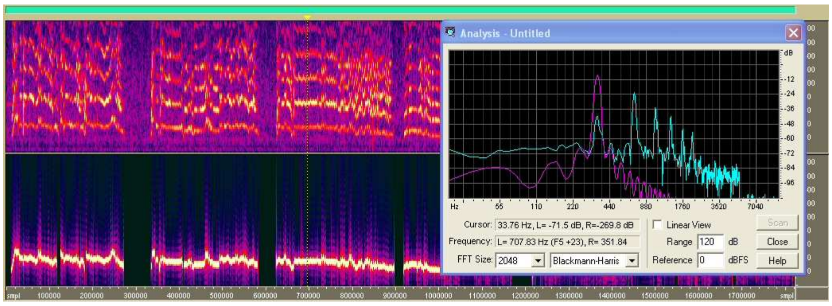
YIN: *de Cheveigne, A. & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. Journal of the Acoustical Society of America, 111 (4), 1917-1930.*

Thanks to authors, we use their shared Matlab implementation

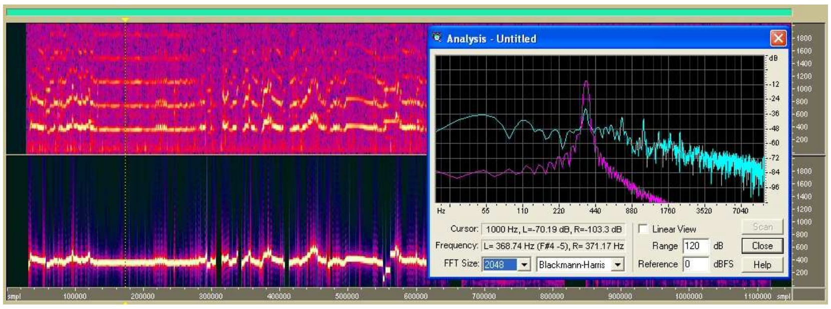
Additional filters: octave correction, filtering out estimations at low energy regions, continuity criteria.



Checking the quality of f0 detection



Cursor: 33.76 Hz, L= -71.5 dB, R=-269.8 dB
Frequency: L= 707.83 Hz (F5 +23), R= 351.84
FFT Size: 2048 Blackmann-Harris
☐ Linear View
Range 120 dB
Reference 0 dBFS



Cursor: 1000 Hz, L=-70.19 dB, R=-103.3 dB
Frequency: L= 368.74 Hz (F#4 -5), R= 371.17 Hz
FFT Size: 2048 Blackmann-Harris
☐ Linear View
Range 120 dB
Reference 0 dBFS

Histogram construction

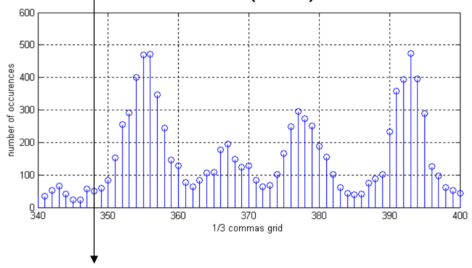
$$Hf_0[n] = \sum_{k=1}^{K} m_k$$

$$m_k = 1,\ f_n \le f_0[k] < f_{n+1}$$

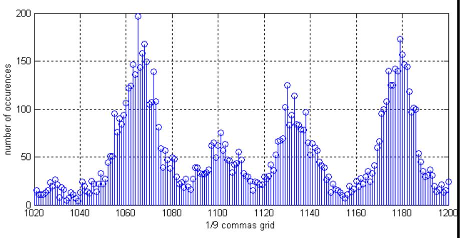
$$m_k = 0, \mathit{otherwise}$$

53Tet: Logarithmic division of an octave into 53 equally spaced intervals.

1/3 Holder comma(53tet) resolution



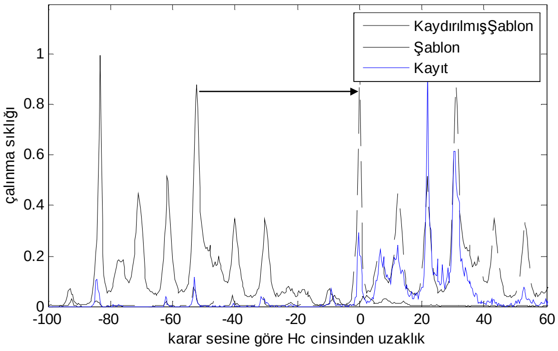
1/9 Holder comma resolution



Resolution of the highest resolution tuning system proposed for Turkish music: 159tet (Yarman, 2008)

A simple idea to perform tonic detection, makam recognition and tuning analysis

Histogram alignment, matching and template construction via cross-correlation



Various distance measures tested for alignment:

L_1-norm: $$d[n]=\frac{1}{K}\sum_{k=0}^{K-1}\left|h_r[k]-h_t[n+k]\right|$$

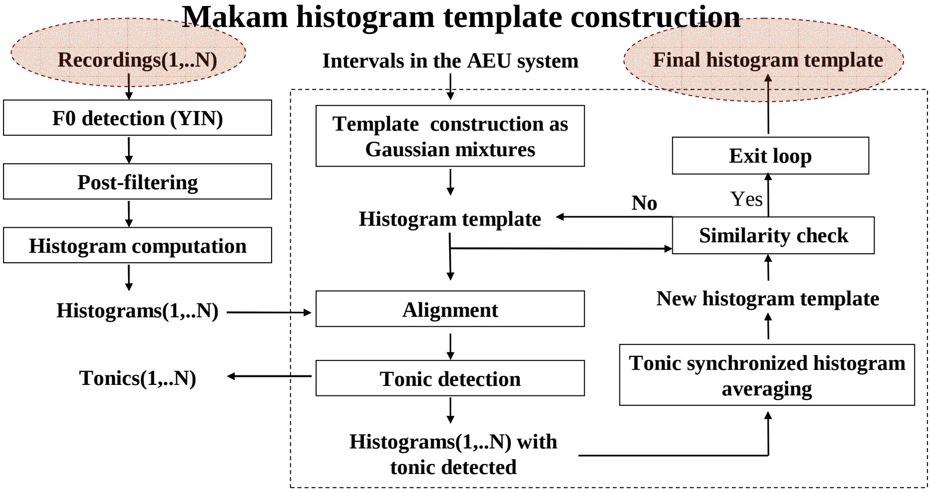
L_2-norm: $$d[n]=\sqrt{\sum_{k=0}^{K-1}\left(h_r[k]-h_t[n+k]\right)^2}$$

Intersection: $$d[n]=\frac{1}{K}\sum_{k=0}^{K-1}\min\left(h_r[k],h_t[n+k]\right)$$

Bhattacharyya distance: $$d[n]=-\log\sum_{k=0}^{K-1}\sqrt{h_r[k]h_t[n+k]}$$

Cross-correlation: $$d[n]=\frac{1}{M}\sum_{k=0}^{K-1}h_r[k]h_t[n+k]$$

Makam histogram template construction



Recordings(1,..N)

F0 detection (YIN)

Post-filtering

Histogram computation

Histograms(1,..N)

Tonics(1,..N)

Intervals in the AEU system

Template construction as Gaussian mixtures

Histogram template

Alignment

Tonic detection

Histograms(1,..N) with tonic detected

Final histogram template

Exit loop

No

Yes

Similarity check

New histogram template

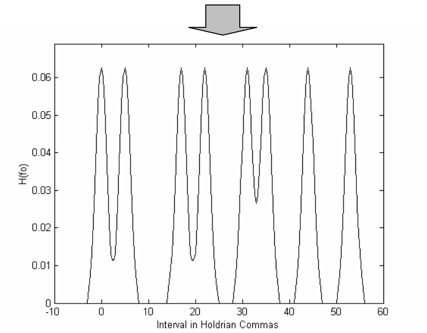
Tonic synchronized histogram averaging

•Bozkurt, B., 2008, "An automatic pitch analysis method for Turkish maqam music," Journal of New Music Research.

•B. Bozkurt, et al, 2009, "Weighing Diverse Theoretical Models On Turkish Maqam Music Against Pitch Measurements: A Comparison Of Peaks Automatically Derived From Frequency Histograms With Proposed Scale Tones", Accepted for publication in Journal of New Music Research.

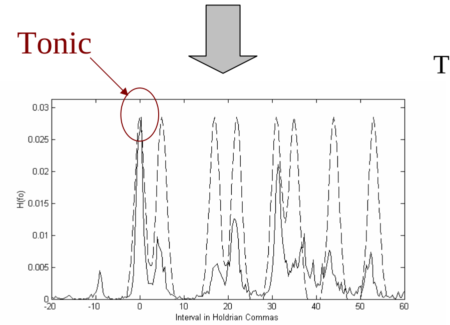
•A. C. Gedik, B. Bozkurt, 2009, "Pitch Frequency Histogram Based Music Information Retrieval for Turkish Music ", Accepted for publication in Signal Processing.

The AEU system scale intervals: [0 5 17 22 31 35 44 53] Hc

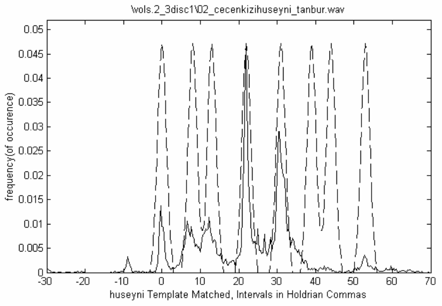


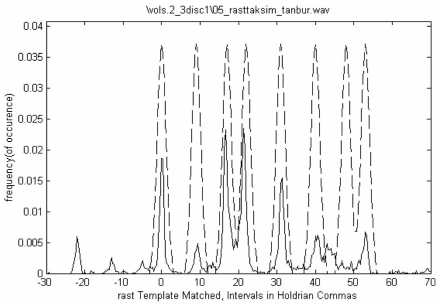
Theoretical *Hicaz* maqam histogram template.

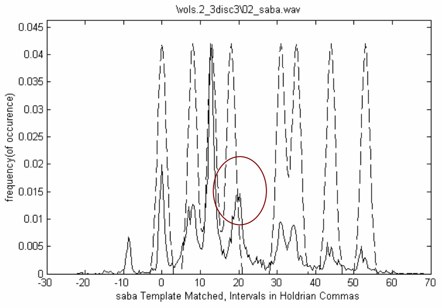
Given a recordings pitch distribution, align the two distributions

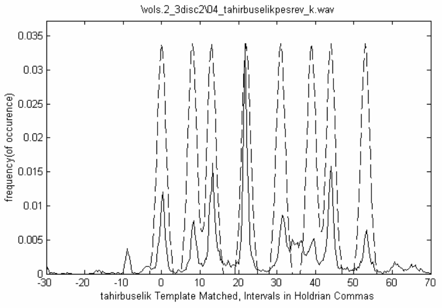


Examples



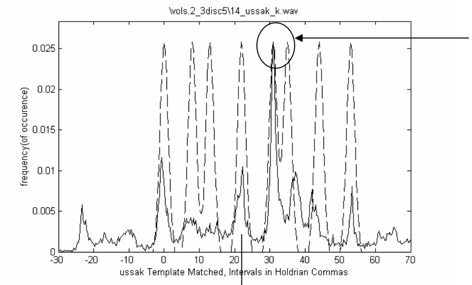




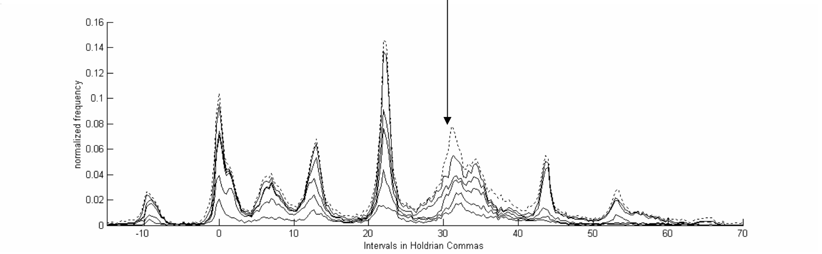


Errors and their contribution in the histogram template

Errors are rather rare (~%5) and have negligible effect on final histogram template

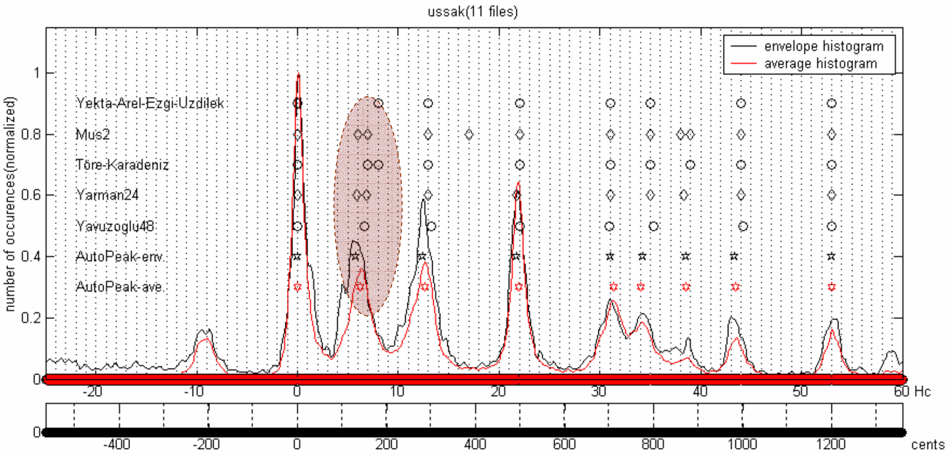


Real tonic



Histogram template for maqam uşşak

Let’s compare our result with the theories



By the way, by automatic tracking of the peaks, we can create automatically a Scala file and play the tuning on a keyboard

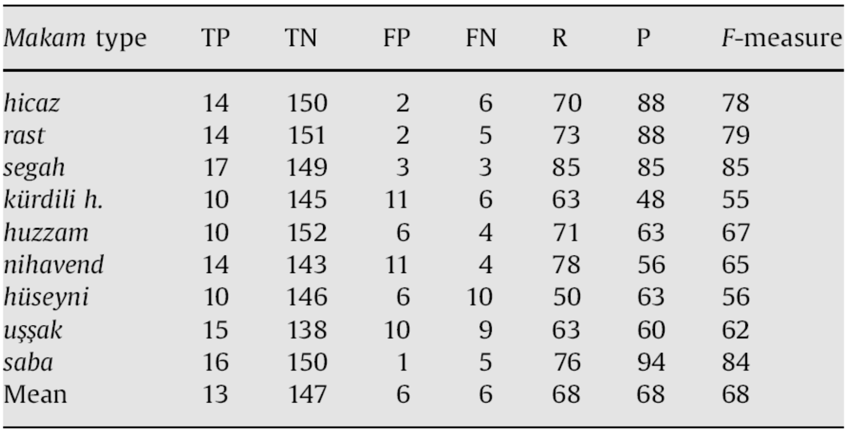
Testing the Tonic Detection on Synthetic Signals (from MIDI)

RAST (24 songs)	CrossCorr.	City Block	Euclidean	Intersection	Bhattacharyya
Mean Tonic F0(Hz)	110.05	109.59	109.51	109.59	109.61
Std (Hz)	**0.42**	**0.16**	**0.19**	**0.16**	**0.17**
MaxDist(Hz)	**1.12**	**0.26**	**0.34**	**0.26**	**0.27**
SEGAH (22 songs)					
Mean Tonic F0(Hz)	276.62	274.71	274.44	274.71	274.16
Std (Hz)	**0.48**	**0.6**	**0.5**	**0.6**	**0.43**
MaxDist(Hz)	**1.18**	**1.27**	**0.95**	**1.27**	**1.23**
Overall (150 songs, 9 makams)					
Mean-Std(Cents)	56.39	2.75	54.62	2.75	2.42
Max-MaxDist(cents)	**579.9**	**7.96**	**583.01**	**7.96**	**7.72**
# false tonic peak det.	**4**	**0**	**4**	**0**	**0**

Tests on 118 real recordings:

# false detections 1 1 1 1 2

Makam recognition via simple template matching



The evaluation results of the *makam* recognition system.

Makam type	TP	TN	FP	FN	R	P	*F*-measure
hicaz	14	150	2	6	70	88	78
rast	14	151	2	5	73	88	79
segah	17	149	3	3	85	85	85
kürdili h.	10	145	11	6	63	48	55
huzzam	10	152	6	4	71	63	67
nihavend	14	143	11	4	78	56	65
hüseyni	10	146	6	10	50	63	56
uşşak	15	138	10	9	63	60	62
saba	16	150	1	5	76	94	84
Mean	13	147	6	6	68	68	68

Distance Mesaure: City-Block(L-1 norm),

Test method: Leave one out

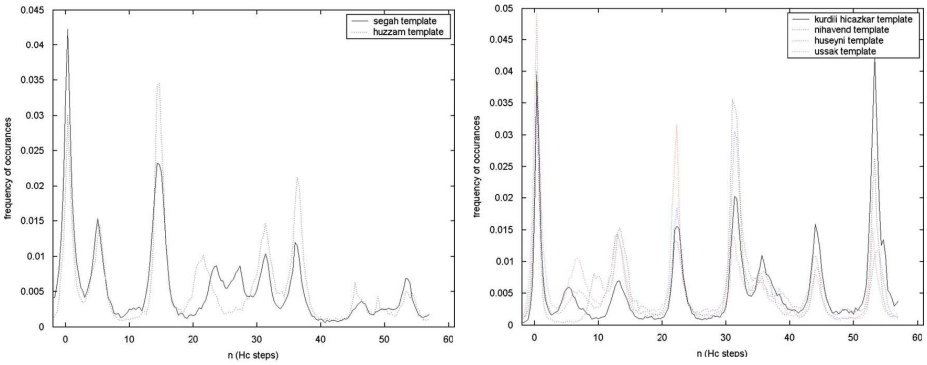
Makam recognition via simple template matching

Confusion matrix

	hicaz	***rast***	***segah***	***kürdili hicazkar***	***hüzzam***	***nihavend***	***hüseyni***	***uşşak***	***saba***
hicaz	**-**	**-**	**-**	**1**	**-**	**2**	**2**	**1**	**-**
rast	**-**	**-**	**-**	**-**	**3**	**1**	**-**	**1**	**-**
segah	**-**	**-**	**-**	**-**	**3**	**-**	**-**	**-**	**-**
kürdili hicazkar	**1**	**-**	**-**	**-**	**-**	**2**	**-**	**3**	**-**
hüzzam	**1**	**-**	**-**	**1**	**-**	**-**	**-**	**-**	**-**
nihavend	**-**	**1**	**1**	**1**	**-**	**-**	**2**	**-**	**-**
hüseyni	**-**	**-**	**-**	**3**	**-**	**3**	**-**	**4**	**-**
uşşak	**-**	**-**	**-**	**4**	**-**	**2**	**2**	**-**	**1**
saba	**-**	**1**	**1**	**1**	**-**	**1**	**2**	**1**	**-**

# of examples: Hicaz:20, Rast:19, Saba:21,Segah:20,Kürdili Hicazkar:16, Hüzzam:14, Nihavend:18, Hüseyni:20,Uşşak:24.

Confused Makams



Highly open to further improvement

Human performance: not measured

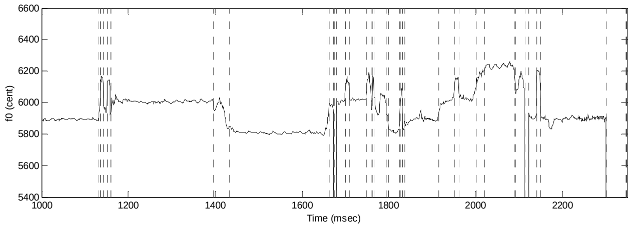
Average Turkish listener should be around 1/9: by chance

Interested Turkish music listener should be around %50 percent

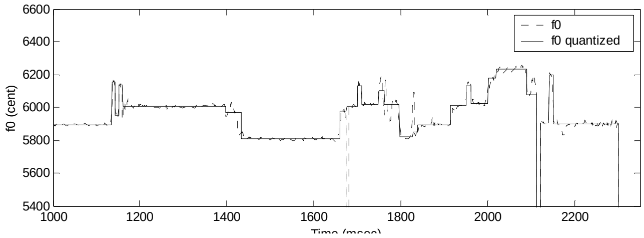
Professional musicians: close to %100

Towards MIDI

Ornamentations need to be labeled first: Heuristic rules applied



Then each chunk is quantized seperately again using rules



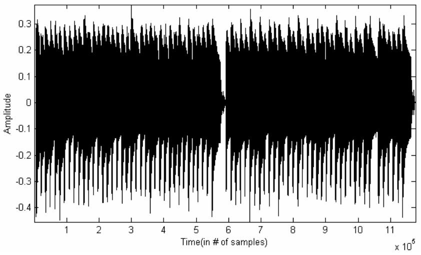
Rhytmic parsing, tempo detection

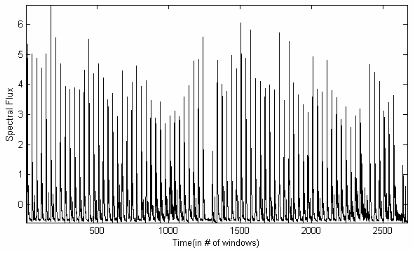
Audio signal

$$SF(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} H(|X(n,k)| - |X(n-1,k)|)$$

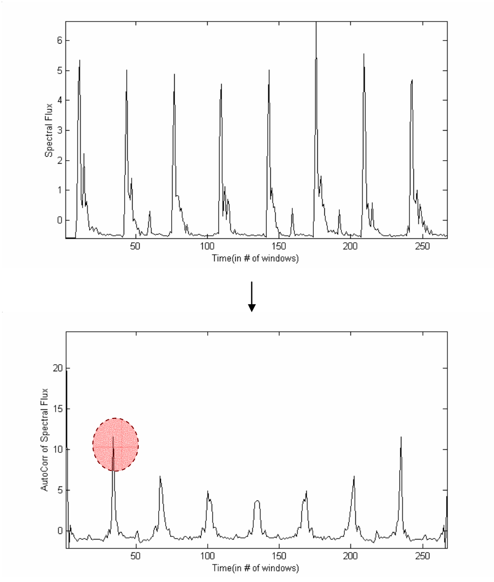
$$H(x) = \frac{x+|x|}{2} \qquad X(n,k) = FT(x_n[k])$$

Spectral Flux





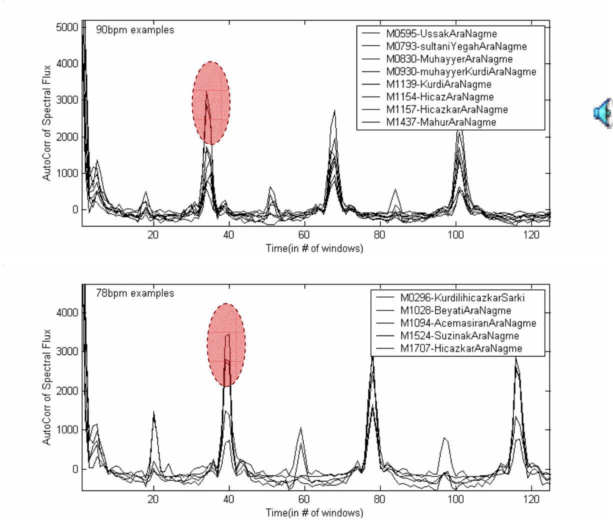
Rhytmic parsing, tempo detection



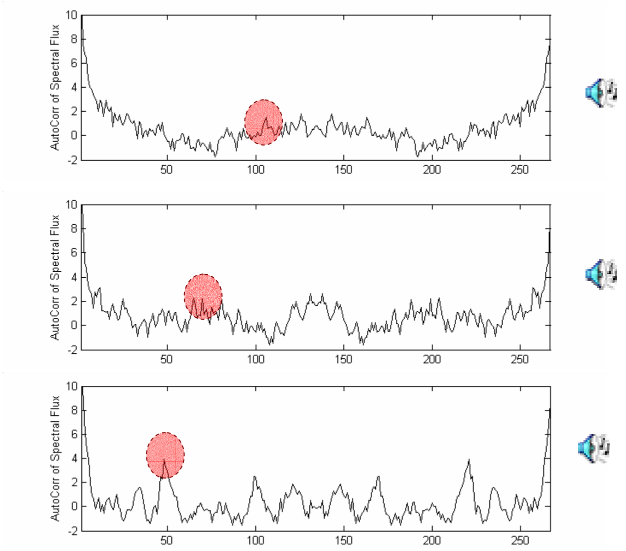
Spectral Flux

Autocorrelation function

Rhytmic parsing, tempo detection
Synthetic audio examples



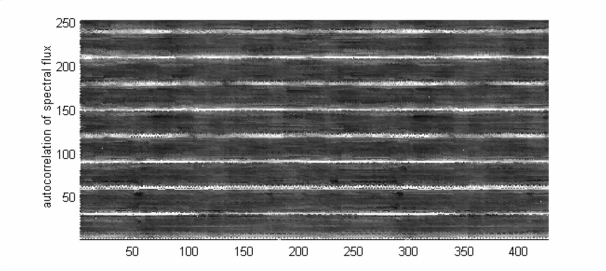
Rhythmic parsing, tempo detection
Real audio examples



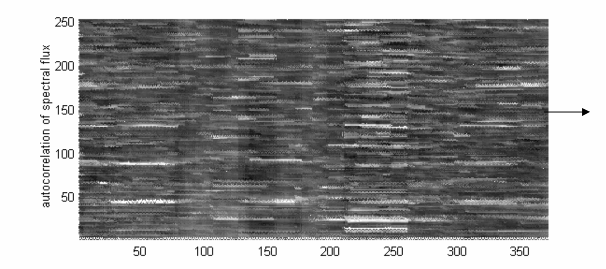
More
percussive

If tempo is varying, we can try something like a spectrogram: The Beat Spectrogram

Synthetic
example

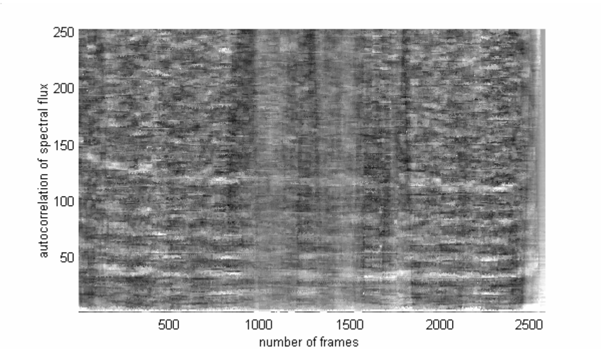


Real
example



→ Hard to
process

The Beat Spectrogram



Begining and end

Can be used as a visual tool but hard to design a robust algorithm to perform automatic detection

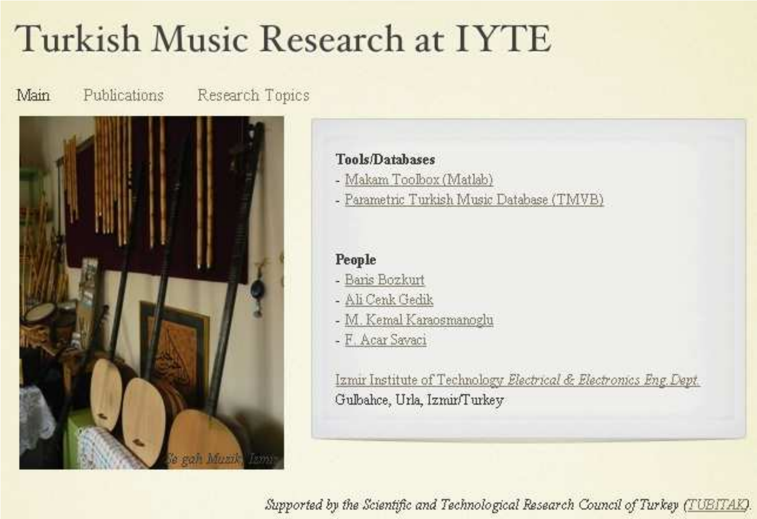
---

Last words

Music technology is not much developed for non-western music, there is a lot of room for research.

Our simple histogram template matching has applications in: Tuner design, instrument standardization, musicology studies on style, historical changes, etc.

THANK YOU FOR YOUR ATTENTION!



Turkish Music Research at IYTE

Main Publications Research Topics

Tools/Databases
- Makam Toolbox (Matlab)
- Parametric Turkish Music Database (TMVB)

People
- Baris Bozkurt
- Ali Cenk Gedik
- M. Kemal Karaosmanoglu
- F. Acar Savaci

Izmir Institute of Technology *Electrical & Electronics Eng.Dept.*
Gulbahce, Urla, Izmir/Turkey

Supported by the Scientific and Technological Research Council of Turkey (TUBITAK).

http://likya.iyte.edu.tr/eee/labs/audio/Main.html